

# 基于改进形状上下文的双焦图像深度估计

冯春<sup>1</sup>, 吴洪涛<sup>1</sup>, 乔兵<sup>2</sup>, 张家驹<sup>1</sup>

(1. 南京航空航天大学机电学院 南京 210016; 2. 南京航空航天大学航天学院 南京 210016)

**【摘要】**针对深度估计时两幅双焦图像特征点的误匹配问题,提出了一种基于改进形状上下文特征点的校正方法。首先引入了对尺度、视角、光照等具有不变性的SSURF特征向量,利用SSURF特征向量匹配算法在双焦单目视觉系统采集的焦距不同的两幅图像之间进行目标SSURF特征点的匹配。提出一种改进的形状上下文描述符,对局部特征匹配点进行误匹配的校正。然后根据空间物点与所成像点距图像中心矢量大小及摄像机的焦距值之间的几何关系,利用校正后的特征点完成相关的计算从而获取目标物的深度信息。实验表明,校正后的方法进行深度估计具有较小的误差和较好的估计效果。该深度估计方法有较大的实用价值。

**关键词** 双焦成像; 深度估计; 改进形状上下文; 误匹配; 单目视觉; SSURF特征点  
中图分类号 TP391 文献标志码 A doi:10.3969/j.issn.1001-0548.2013.02.017

## Depth Estimation for the Bifocal Image Based on Improved Shape Context

FENG Chun<sup>1</sup>, WU Hong-tao<sup>1</sup>, QIAO Bing<sup>2</sup>, and ZHANG Jia-ju<sup>1</sup>

(1. School of Mechanical Electrical, Nanjing University of Aeronautics and Astronautics Nanjing 210016;  
2. School of Astronautics, Nanjing University of Aeronautics and Astronautics Nanjing 210016)

**Abstract** A correction method for feature points mismatching is proposed based on improved shape context, aiming at the mismatching problem of corresponding points needed for the depth estimation for the bifocal image. First, the SSURF feature descriptor is presented, as it has good invariability to changes such as scaling, view angles and illumination. By the use of SSURF feature vector matching, object images which had been obtained based on a monocular stereo vision system are processed in two different focal length images, so corresponding SSURF feature points are obtained. Then, an improved shape context descriptor is proposed to correct the error matching of local feature points. The distance between dimensional target points as well as image points and the center of the image, and the geometric relationship between the camera focal lengths are used to complete the related computation and therefore to obtain the depth information of object by using the corrected feature points. The experimental results show that the correction method introduced has small errors for depth estimation and good effect for estimation.

**Key words** bifocal imaging; depth estimation; improved shape context; mismatching; monocular vision; SSURF feature points

计算机视觉的核心问题之一是从物体的2D图像恢复3D信息,为此学者们提出了许多方法,其中最常用的是立体视觉方法和运动视觉方法。单目立体视觉由于只需使用一台图像采集设备获取场景的2D图像信息,并通过单目立体视觉算法恢复场景的3D信息,因而成为计算机视觉领域一个重要的研究分支。单目立体视觉算法大致分为基于对焦<sup>[1]</sup>算法、基于离焦<sup>[2]</sup>算法和基于变焦<sup>[3-4]</sup>算法3种。对焦算法要求多幅图像,对硬件条件要求较高,实时性差,但

是精度高;离焦算法需要两幅或者多幅图像,缺点是点扩散函数计算复杂,深度图精度低;变焦算法也需要两幅或者多幅图像,与离焦算法相比其所需的摄像机参数较少并且容易获取,算法的复杂度低。双焦算法<sup>[5]</sup>本质上可以看作是一种更为简洁的变焦算法,它通过改变焦距对空间物体拍摄得到两幅清晰而物像大小不同的图像,根据空间物点与其所成像点距图像中心的矢量大小及摄像机的焦距值之间的几何关系,完成相关的处理得到物体的深度信息。

收稿日期: 2011-06-14; 修回日期: 2012-05-07

基金项目: 国家自然科学基金(51075209)

作者简介: 冯春(1981-),男,博士生,主要从事机器人控制、视觉导航方面的研究。

图像深度是图像理解<sup>[6-8]</sup>、三维立体重建、机器人视觉<sup>[9]</sup>等研究的前提和基础。

本文利用单目双焦图像进行深度估计, 主要针对深度估计所需两幅图像对应点之间误匹配问题做深入的研究。考虑到双焦图像存在尺度变化, 故首先采用基于SSURF描述子进行图像匹配获取目标特征匹配点对, 并提出一种改进的形状结构上下文描述符进行局部特征误匹配的校正, 以保证匹配的正确率。然后根据空间物点与所成像点距图像中心矢量大小及摄像机的焦距值之间的几何关系, 利用校正后的特征点完成相关的计算从而更准确获取目标物的深度信息。

## 1 基于局部特征点的图像匹配

SURF算法是对SIFT<sup>[10]</sup>算法的一种改进, 主要思想是降低特征描述子维数, 以减少匹配时间, SURF算法可参阅文献<sup>[11]</sup>。

### 1.1 SSURF算子

在SURF算法中, 为了保持旋转不变, 要求兴趣点满足可重复的方向。因此, 需要在某一圆形区域邻域内计算利用空间向量表示的哈尔小波响应, 通过水平和垂直哈尔小波响应及一个滑动方向窗口获取主方向。小尺寸窗口对应单一的主梯度, 大尺寸窗口则生成向量长度极大值。它们均导致了在兴趣区域的定向误差。

基于双焦单目视觉的图像之间并不发生旋转变换, 故可以将上述步骤略去生成SSURF算子, 以此来提高算法的速度和获取更高的实用性, 相关内容可参考文献<sup>[11]</sup>。

### 1.2 特征点的匹配

基于SSURF算法进行图像特征点提取后采用最近邻域匹配法, 找出所有的可能匹配点对。

实验发现由于局部特征描述符仅利用了特征点的局部信息, 当图像中包含一定程度相似或对称结构时(普遍存在), 容易发生误匹配。通过实验, 误匹配主要有两种: 1) “单点对多点”误匹配, 即参考点集合的某点匹配待检测点集合的多点; 2) “单点对单点”误匹配, 即参考点集合的某点匹配待检测点集合的某点。

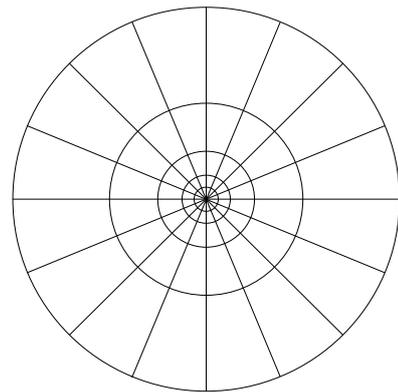
## 2 基于改进形状上下文的局部特征误匹配校正

为了准确进行图像的深度信息估计, 要求特征点的匹配必须准确, 否则获取的深度信息误差很大。

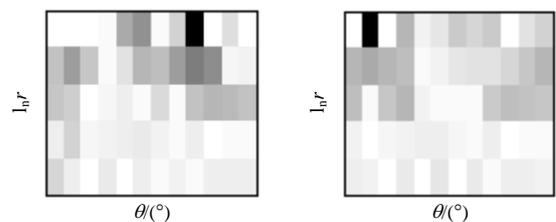
考虑到基于局部特征描述符的图像匹配利用的是局部特征信息, 并未考虑特征点的空间拓扑结构, 在完成图像匹配尤其是具有大量相似或者对称结构的图像匹配情况时, 容易出现误匹配。故本文提出一种改进的形状上下文描述符对局部特征误匹配进行校正。

### 2.1 改进的形状上下文

形状上下文算法<sup>[12]</sup>的主要思想是提取图像目标轮廓上的若干控制点, 建立对数极坐标系, 通过计算这些控制点之间的距离以及角度信息完成目标物的形状匹配。改进的形状上下文描述符是一种结合了形状上下文描述符与局部特征描述符优点的新的描述符。它首先利用局部特征描述符得到特征点, 再利用形状上下文的思想, 计算特征点的拓扑结构关系。由于可以描述物体各结构之间的大致空间关系, 所以能有效匹配出那些具有类似或对称结构的目标物。



a. 改进形状上下文对数极坐标



b. 某两个局部特征点的改进形状上下文

图1 改进形状上下文

对于特征点的提取, 本文不作介绍, 具体的做法如SURF算法所述。假设得到了一幅图像的局部特征点, 则改进描述符按如下步骤计算。

首先利用特征点之间的平均值距离对每个特征点对之间的距离进行正则化, 以保证图像的尺度不变性; 然后为每一个特征点建立一个对数极坐标系, 如图1所示, 假定构造一个 $5 \times 12$ 的坐标系, 并且这个对数极坐标系统的半径分别设置为 $r$ 、 $r/2$ 、

$r/4$ 、 $r/8$ 、 $r/16$ ，此处 $r$ 取值为2。这样只需要统计每个格子内特征点的尺度值就能够得到每个特征点的改进描述符。它实际上是直方图的表示，如图1所示，可以看到每一行代表一个半径内的特征点分布情况，而每一列表示某个方向上特征点的分布情况。为了避免尺度值小的特征点和尺度值大的特征点进行匹配(往往是错误匹配)，需要将每个特征点的改进描述符乘以一个系数，即该特征点的尺度值和最大尺度值的比值来进行校正。

每一特征点改进描述符的修正结果为：

$$h_i(k) = \sum \left\{ \frac{s(p_i) \times s(p_j)}{\max\_s} \mid p_j \in \text{bin}_i(k) \right\} \quad (1)$$

式中， $p_i$ 代表特征点； $s(p_i)$ 是特征点 $p_i$ 的尺度值； $\max\_s$ 是图像全部特征点的最大尺度值。尺度值可以定义为图1a所示极坐标中各格子内边缘关键点的个数。具体做法是首先通过Canny等边缘算法提取图像的边缘点，对其进行采样，定义采样点为图像的边缘关键点；然后对每一个局部特征点构造一个图1a所示的对数极坐标，计算每一个格子内的边缘关键点的个数即为尺度值。考虑到图像的复杂程度不同，特征点的数量会有变化，故需要将上述改进的描述符再次进行归一化，以此得到对于特征点之间空间拓扑结构关系更有效的描述符。

## 2.2 两种校正匹配形式

对于误匹配点对的确定，可以利用相似度的准则进行判断。本文采用基于特征变换的方法进行相似度计算。提出的两种匹配校正算法具体的实现过程为：取参考点集内的某一特征点，判断待匹配点集中是否有多个特征点与其匹配，若存在“一对多”匹配则必定存在误匹配；否则为“一对一”匹配，利用相似度准则进行判断是否误匹配。

### 2.2.1 “单点对多点”误匹配校正

首先介绍相似度的概念，假定存在两个点 $a$ 、 $b$ ，则它们的描述子为 $D_a$ 、 $D_b$ ，假定其向量维数为 $l_a$ 、 $l_b$ ，相似度定义<sup>[13]</sup>为：

$$S = \frac{1}{\frac{1}{M} \sum_{j=1}^M |D_a(j) - D_b(j)|} \quad (2)$$

式中， $M = \max(l_a, l_b)$ ； $D_a(j)$ 、 $D_b(j)$ 分别表示改进描述符的分量。

对于“单点对多点”的误匹配，校正误匹配的过程如下：假设存在两个匹配点集，定义为

$A = \{a_1, a_2, \dots, a_m\}$ 和 $B = \{b_1, b_2, \dots, b_n\}$ ，为了描述方便，本文以如图2所示的“单点对两点”的误匹配为例进行说明。

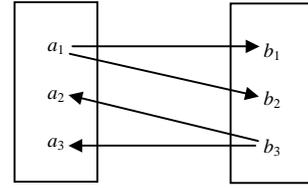


图2 “单点对两点”误匹配

首先以点集合 $A$ 为参考点集，分别计算点 $a_1$ 与点 $b_1$ 的改进描述符 $D_{a_1}$ 与 $D_{b_1}$ ，以及点 $a_1$ 与点 $b_2$ 对应的改进描述符 $D_{a_1}$ 与 $D_{b_2}$ ，并依据式(2)计算两个匹配点对的相似度，取相似度大的作为正确的匹配点对保留。再以点集合 $B$ 为参考点集，计算点 $b_3$ 与点 $a_2$ 对应的改进描述符 $D_{b_3}$ 与 $D_{a_2}$ ，以及点 $b_3$ 与点 $a_3$ 对应的改进描述符 $D_{b_3}$ 与 $D_{a_3}$ 并计算两个匹配点对的相似度，并选取正确的匹配点对。

### 2.2.2 “单点对单点”误匹配校正

通过计算改进描述符对其余“单点对单点”的局部特征匹配点对进行判断，如果确定为误匹配，则剔除误匹配点对。具体的判定方法如下。

假定有一匹配点对 $(a_i, b_i)$ ，计算得到其改进描述符分别为 $D_1(i)$ 、 $D_2(i)$ ，其向量长度为 $l_1(i)$ 、 $l_2(i)$ ，首先将求得的描述符进行差运算，得到：

$$d(i) = |D_1(i) - D_2(i)| \quad (3)$$

式中， $i = 1, 2, \dots, n$ ， $n = \max(l_1, l_2)$ 。计算 $d(i) > T_1$ 的个数 $N$ ，其中， $T_1$ 为阈值，表示与两个匹配点对应的图像区域的特征点数目的门限值，通常选取30~50。然后计算比值 $r = \frac{N}{\max(l_1, l_2)}$ ，若 $r < T_2$ ，则确定

局部特征匹配点对为正确匹配；反之，则确定为错误匹配，去除掉。其中， $T_2$ 是匹配阈值，通常设定为0.2~0.5。

## 3 双焦单目视觉深度估计

双焦成像算法<sup>[5]</sup>的基本原理为：在两幅变焦图像中找到匹配点对，获取所成像的矢量位移的长度并给出焦距值，利用空间物点深度与其在不同焦距下所成像的矢量位移与相应焦距大小之间的几何关系，从而计算出空间物点深度信息。依据图3所示几何模型的关系可以推导出深度为：

$$Z = \frac{r_2 f_1 (f_1 - f_2)}{r_1 f_2 - r_2 f_1} = -(f_1 + f_2 \frac{1}{ar_1/\Delta r - 1}) \quad (4)$$

式中,  $Z$  为深度;  $f_1$ 、 $f_2$  分别代表较小和较大焦距值;  $r_1$ 、 $r_2$  分别为相应的像点矢量大小;  $\Delta r = r_2 - r_1$ ;  $a = (f_2 - f_1)/f_1$ 。由上式可知, 当双焦相机的焦距  $f_1$ 、 $f_2$  已知时, 深度  $Z$  可由  $ar_1/\Delta r$  唯一确定。

假定两幅图像中对应点的坐标分别为  $(X_1, Y_1)$  和  $(X_0, Y_0)$ , 则通过简化三维模型得到两幅图像中对应点有如下二维运动关系:

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} = S \begin{bmatrix} 1 & -\theta \\ \theta & 1 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (5)$$

式中,  $S$  为变焦系数;  $\theta$  为绕  $Z$  轴的旋转角;  $(\Delta x, \Delta y)$  为两幅图像的平移矢量。故需要在两幅图像中利用局部特征图像匹配提取  $N$  对特征点并进行特征点匹配, 建立方程组, 利用最小二乘法计算上述 4 个参数, 再利用摄像机标定的内参数, 完成深度信息估计。

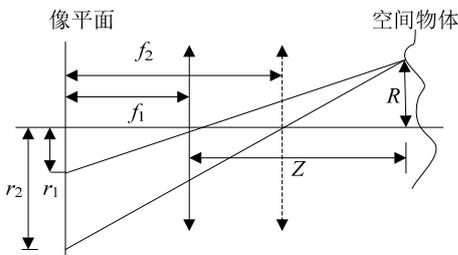


图3 成像几何模型

### 4 目标物深度估计实验

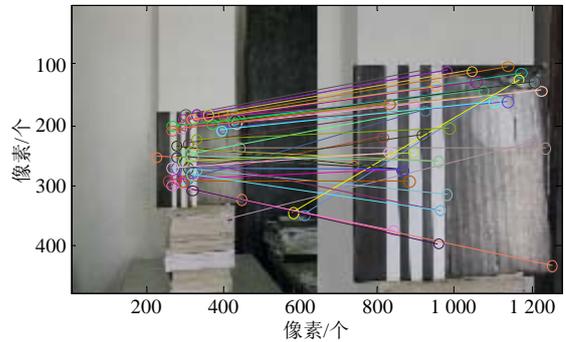
实验中利用 Canon PowerShot A2000 IS 型彩色数码相机采集图像, 本文利用手动摄像图片完成实验, 采集图像的大小为  $640 \times 480$  像素, CCD 面积约  $5.56 \text{ mm} \times 4.17 \text{ mm}$ , 焦距  $f_1$  取最小焦距, 焦距  $f_2$  取 2.3X, 选取的大小标称焦距值分别为  $6.4 \text{ mm}$  和  $14.72 \text{ mm}$ , 最大光圈 F3.2/5.9, 由于深度估计是基于  $35 \text{ mm}$  胶片相机得出的, 所以在估算深度时应考虑折算倍率的修正。本文分别以纸盒和墙壁上的插座作为目标物完成实验。

#### 4.1 实验1

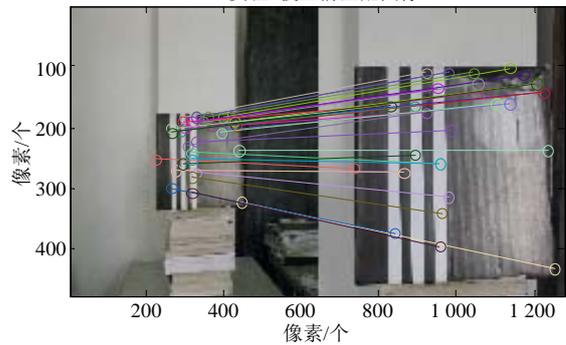
首先固定相机与纸盒的距离(约  $500 \text{ mm}$ ), 在焦距  $f_1$  和  $f_2$  下分别获取清晰图像, 并调整变焦系数 ( $0.3 \sim 0.7$ ) 进行多组实验。实验过程如下: 利用 SSURF 局部特征描述符进行图像匹配, 匹配结果如图 4a 所示。利用本文提出的改进形状上下文描述符对误匹配进行校正, 结果如图 4b 所示。

为了进行比较, 分别利用校正前后的特征点进

行深度计算, 结果如表 1、表 2 所示。由表 1 可知, 校正之前估算的深度值误差较大, 最小误差值约为  $52 \text{ mm}$ ; 由表 2 可知, 校正后的估算深度值最大误差约为  $22 \text{ mm}$ , 改进后的深度估计值精度大大提高。



a. 实验1校正前匹配图像



b. 实验1校正后匹配图像

图4 实验1匹配图像

表1 实验1校正前结果

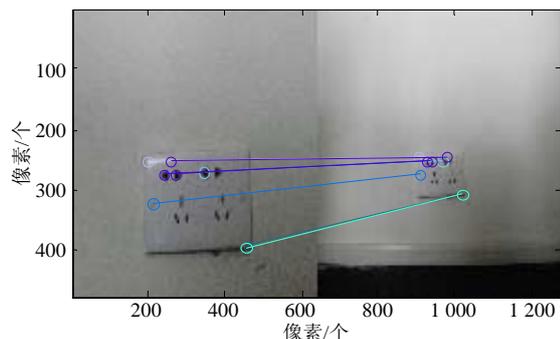
序号	校正前		
	匹配点对数目	正确匹配数目	估算深度/mm
1	54	35	420.2
2	31	21	448.1
3	20	11	553.6
4	49	38	423.5
5	29	22	562.8
6	23	14	441.1

表2 实验1校正后结果

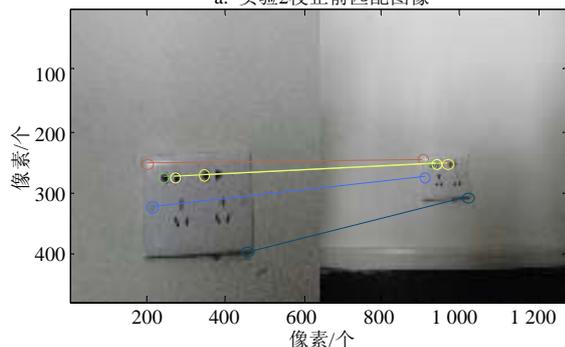
序号	校正后		
	匹配点对数目	正确匹配数目	估算深度/mm
1	36	34	486.9
2	21	21	513.5
3	11	11	479.6
4	38	37	521.7
5	22	20	518.3
6	14	14	488.1

#### 4.2 实验2

与实验 1 实验条件相似, 插座目标校正前和校正后的结果如图 5a、图 5b 所示, 结果比较如表 3、表 4 所示。



a. 实验2校正前匹配图像



b. 实验2校正后匹配图像

图5 实验2匹配图像

表3 实验2校正前结果

序号	校正前		
	匹配点对数目	正确匹配数目	估算深度/mm
1	11	5	422.3
2	6	5	519.8
3	3	3	490.1
4	11	6	451.5
5	10	4	435.6
6	15	6	417.3

表4 实验2校正后结果

序号	校正后		
	匹配点对数目	正确匹配数目	估算深度/mm
1	5	5	510.8
2	5	5	513.5
3	3	3	493.6
4	6	6	486.9
5	4	4	488.5
6	6	6	513.1

4.3 实验3

前两组实验是固定物距，改变变焦系数估算同一深度值，以验证本文提出的深度估计方法的有效性。实验3是在实验1的实验条件下，固定变焦系数(取0.5)，改变物距即深度值进行的实验，表5给出校正前后的估算深度。可以看到，校正后的精度比校正前大大提高。

表5 实验3结果 单位: mm

序号	校正前	校正后	测量值(近似值)
1	519.3	588.6	600
2	622.9	710.1	700
3	855.7	779.2	800
4	935.8	986.9	1 000

实验1~3校正后的深度值，与文献[5]相比，本

文的方法减小了深度估计误差。分析其误差的原因主要有：1) 摄像机标定的内参数与实际内参数存在偏差；2) 摄像机的分辨率大小不同。

4.4 实验4

为了进一步验证本文提出的改进形状上下文描述符对局部特征误匹配校正的有效性，单独采用2.2节的校正方法对局部特征描述符的误匹配进行校正并与上述实验结果进行对比，限于篇幅只给出实验1、实验2条件下的校正及深度估计结果，如表6、表7所示。

表6 实验1条件下的校正结果

序号	校正后		
	匹配点对数目	正确匹配数目	估算深度/mm
1	40	34	444.2
2	25	20	572.3
3	13	9	460.9
4	43	35	568.2
5	26	21	455.1
6	18	12	463.7

表7 实验2条件下的校正结果

序号	校正后		
	匹配点对数目	正确匹配数目	估算深度/mm
1	7	5	483.7
2	5	5	490.6
3	3	3	518.8
4	8	6	484.4
5	7	4	520.3
6	9	6	535.4

由实验4结果可知：与表1、表3相比，该实验同样减小了深度估计误差，分析其原因是采用的误匹配校正方法将“一对多”误匹配校正为“一对一”的匹配形式，从而消除了一部分误匹配点对；与表2、表4相比，实验深度估计误差较大，原因是虽然误匹配校正方法可以消除“一对多”的误匹配，但是仍存在部分“一对一”的误匹配，其本质原因仍是局部特征描述符的“局部性”，单独采用相似度准则不能很好避免“一对一”误匹配；改进形状上下文可以进一步改善误匹配的校正效果，从而提高深度估计精度。

变焦系数是本文考虑的实验条件之一，实验1、2、4对应的变焦系数分别为0.3、0.5和0.7。变焦系数与深度估计性能存在一定关系，由文献[5]中双焦深度估计原理可知，在相机标定及图像匹配条件一致的情况下，变焦系数越小则深度估计鲁棒性越好。然而变焦系数越小则两幅图像越容易出现误匹配(不考虑图像背景的影响)，故变焦系数应在一定范围内选取，本文取0.3~0.7。

## 5 结 论

利用SSURF特征匹配算法在不同焦距的两幅图像中提取特征匹配点对集合, 采用改进的形状上下文描述符进行特征点之间误匹配校正, 获取特征点集合。在此基础上基于双焦单目立体视觉通过几何关系能够更准确获取目标物的深度信息。

本文的双焦单目视觉目标深度估计实验, 分别以纸盒和墙壁上的插座为目标物, 依次基于固定物距和改变焦比系数的条件下完成深度估算。实验结果证明计算所得的深度与实测结果基本符合, 利用该方法进行深度估计具有较高的精度, 较好的适应性, 可用于在特定背景下进行目标物的深度信息估计, 具有一定的实用价值。本文下一步的工作是在所提出的深度估计方法的基础上, 获取稠密深度图, 进一步完成图像3D重建等任务。

### 参 考 文 献

- [1] MA Jun, OLSEN S I. Depth from zooming[J]. Journal of the Optical Society of American A(Optics and Images Science), 1990, 7(10): 1883-1890.
- [2] XIONG Ya-lin, SHAFER S A. Depth from focusing and defocusing[C]//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 1993: 68-73.
- [3] KROTKOV E. Focusing[J]. International Journal of Computer Vision, 1988, 1(3): 223-237.
- [4] ASADA N, BABA M, ODA A. Depth from blur by zooming [C]//Proceedings of the Vision Interface Annual Conference. Ottawa, Canada: [s.n.], 2001: 165-172.
- [5] 王剑, 王元庆. 基于双焦成像的单眼立体视觉算法[J]. 机器人, 2007, 29(1): 41-44.  
WANG Jian, WANG Yuan-qing. A monocular stereo vision algorithm based on bifocal imaging[J]. Robot, 2007, 29(1): 41-44.
- [6] 狄红卫, 柴颖, 李逵. 一种快速双目视觉立体匹配算法[J]. 光学学报, 2009, 29(8): 2180-2184.  
DI Hong-wei, CHAI Ying, LI Kui. A fast binocular vision stereo matching algorithm [J]. Acta Optica Sinica, 2009, 29(8): 2180-2184.
- [7] 赵钦君, 赵东标, 韦虎. Harris-SIFT算法及其在双目立体视觉中的应用[J]. 电子科技大学学报, 2010, 39(4): 546-550.  
ZHAO Qin-jun, ZHAO Dong-biao, WEI Hu. Harris-SIFT algorithm and its application in binocular stereo vision [J]. Journal of University of Electronic Science and Technology of China, 2010, 39(4): 546-550.
- [8] 王传宇, 赵明, 阎建河, 等. 基于双目立体视觉技术的玉米叶片三维重建[J]. 农业工程学报, 2010, 6(4): 198-202.  
WANG Chuan-yu, ZHAO Ming, YAN Jian-he, et al. Three-dimensional reconstruction of maize leaves based on binocular stereovision system[J]. Transactions of the CSAE, 2010, 6(4): 198-202.
- [9] 张磊, 王书茂. 基于双目视觉的农田障碍物检测[J]. 中国农业大学学报, 2007, 12(4): 70-74.  
ZHANG Lei, WANG Shu-mao. Detection of obstacles in farmland based on binocular vision[J]. Journal of China Agricultural University, 2007, 12(4): 70-74.
- [10] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [11] BAY H, TUYTELAARS T, GOOL L V. SURF: Speeded up robust features[C]//Proceedings Asian Conference on Computer Vision. [S.l.]: [s.n.], 2006, 1: 918-927.
- [12] BELONGIE S, MALIK J, PUZICHA J. Shape matching and object recognition using shape context[J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2002(2): 509-522.
- [13] 周咏梅, 韩国强. 一种基于区域的图像相似性计算方法 [J]. 广西师范大学学报: 自然科学版, 2008, 26(1): 224-227.  
ZHOU Yong-mei, HAN Guo-qiang. A region-based image similarity computation approach[J]. Journal of Guangxi Normal University(Natural Science Edition), 2008, 26(1): 224-227.

编辑 漆蓉